

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-181485
(P2000-181485A)

(43) 公開日 平成12年6月30日 (2000.6.30)

(51) Int.Cl. ⁷	識別記号	F I	キーワード (参考)	
G 1 0 L 15/22		G 1 0 L 3/00	5 7 1 V	2 C 0 3 2
	15/06		5 2 1 W	2 F 0 2 9
	15/00		5 5 1 Q	5 D 0 1 5
G 0 1 C 21/00		G 0 1 C 21/00	H	5 H 1 8 0
G 0 8 G 1/0969		G 0 8 G 1/0969		

審査請求 未請求 請求項の数 4 O L (全 6 頁) 最終頁に続く

(21) 出願番号 特願平10-354995

(22) 出願日 平成10年12月14日 (1998.12.14)

(71) 出願人 000003207

トヨタ自動車株式会社
愛知県豊田市トヨタ町1番地

(72) 発明者 市原 雅明

愛知県豊田市トヨタ町1番地 トヨタ自動車株式会社内

(74) 代理人 100075258

弁理士 吉田 研二 (外2名)

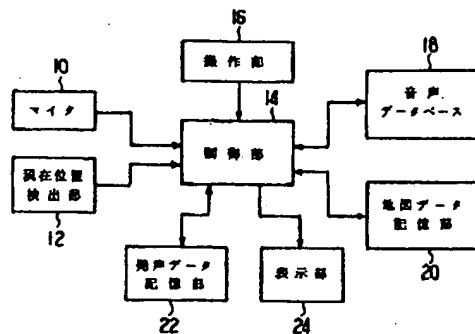
最終頁に続く

(54) 【発明の名称】 音声認識装置及び方法

(57) 【要約】

【課題】 音声でナビゲーション用の目的地を設定する際に、ユーザの発声負担を軽減する。

【解決手段】 ユーザはマイク10からナビゲーション用の目的地データを入力する。制御部14は、入力発声データを発声データ記憶部22に記憶するとともに、音声データベース18を用いて入力発声データの少なくとも一部を解析する。解析して得られた結果を用いて解析に使用する音声データベース18を切替え、発声データ記憶部22に記憶された発声データを読み出して再解析を行う。記憶された発声データを用いるので、ユーザは1回の発声だけで高精度に目的地を設定できる。



【特許請求の範囲】

【請求項1】 ユーザの発声データを記憶する発声データ記憶手段と、

前記発声データと音声データベース内の音声データとを比較することにより前記発声データの少なくとも一部を解析する第1音声解析手段と、

前記第1音声解析手段で得られた解析データに基づいて前記音声データベースを切り替える切替手段と、

前記発声データ記憶手段に記憶された発声データを読み出し、前記切替手段で切り替えられた音声データベース内の音声データとを比較することにより前記発声データを再解析する第2音声解析手段と、

を有することを特徴とする音声認識装置。

【請求項2】 請求項1記載の装置において、前記発声データはナビゲーション用の目的地データであり、

前記第2音声解析手段で所定のデータが得られた場合に、該データを前記目的地用の目標物として処理する手段を有することを特徴とする音声認識装置。

【請求項3】 ユーザの発声データを記憶する記憶ステップと、

前記ユーザの発声データと音声データベース内の音声データとを比較することにより前記音声データの少なくとも一部を解析する第1解析ステップと、

前記第1解析ステップで得られた解析データに基づいて前記音声データベースを切り替える切替ステップと、

前記記憶ステップで記憶された発声データを読み出し、読み出された発声データと前記切替ステップで切り替えられた音声データベース内の音声データとを比較することにより再解析する第2解析ステップと、

を有することを特徴とする音声認識方法。

【請求項4】 請求項3の方法において、前記発声データはナビゲーション用の目的地データであり、前記第2解析ステップにおいて所定のデータが得られた場合に、該データを前記目的地用の目標物として処理する処理ステップをさらに有することを特徴とする音声認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は音声認識装置及び方法、特にナビゲーションシステムにおいて目的地を設定する際の音声認識に関する。

【0002】

【従来の技術】 従来より、音声でナビゲーションシステムの各種処理、例えば目的地設定を行う技術が提案されている。このような技術においては、いかに迅速かつ正確にユーザが発した音声を認識するかが重要な課題である。通常、音声認識はユーザの発声データと予め用意された音声データベース内の音声データとを比較すること

で行われるが、音声データベースは階層化して用いられることが多い。

【0003】 例えば、特開平10-62199号公報には、音声データベースを3つの階層に分け、階層1に位置情報を有する施設名と位置情報を有しない施設ジャンル名を記憶し、階層2に階層1のジャンル名に該当する、位置情報を有する施設名と位置情報を有しない都道府県名を記憶し、階層3に階層2の都道府県に該当する、位置情報を有する施設名を記憶して、ユーザの発声データに応じて順次階層を変化させて音声認識する技術が開示されている。

【0004】

【発明が解決しようとする課題】 しかしながら、上記従来技術ではユーザの発声毎に音声データベースの階層を切り替えているため、例えばユーザがナビゲーションの目的地をデパートの**百貨店に設定したいと欲しても、順次「施設」→「デパート」→「**百貨店」と繰り返し発声しなければならず、「デパートの**百貨店に行きたい」などのように自然な発声1回で目的地を設定することができなかった。

【0005】 また、ユーザによっては、例えば「**百貨店の近くの駐車場」の如く、ある目標物を起点として目的地を設定したいと欲する場合があるが、従来技術ではこのような目標物を起点とした目的地設定を認識することができない問題があった。

【0006】 本発明は、上記従来技術の有する課題に鑑みなされたものであり、その目的は、ユーザの発声の負担を軽減してより簡易に所望のデータを音声で設定できる装置及び方法を提供することにある。

【0007】

【課題を解決するための手段】 上記目的を達成するために、第1の発明は、ユーザの発声データを記憶する発声データ記憶手段と、前記発声データと音声データベース内の音声データとを比較することにより前記発声データの少なくとも一部を解析する第1音声解析手段と、前記第1音声解析手段で得られた解析データに基づいて前記音声データベースを切り替える切替手段と、前記発声データ記憶手段に記憶された発声データを読み出し、前記切替手段で切り替えられた音声データベース内の音声データとを比較することにより前記発声データを再解析する第2音声解析手段とを有することを特徴とする。第1音声解析手段で解析し、さらに記憶しておいた発声データを読み出して第2音声解析手段で再解析する（再解析時には、音声データベースが切り替えられて最適化されている）ことで、ユーザの1回の発声で確実に音声認識を行うことができる。なお、第1音声解析手段と第2音声解析手段は別個に存在する必要はなく、同一の手段で両機能を達成することも可能である。

【0008】 また、第2の発明は、第1の発明において、前記発声データはナビゲーション用の目的地データ

であり、前記第2音声解析手段で所定のデータが得られた場合に、該データを前記目的地用の目標物として処理する手段を有することを特徴とする。音声解析して得られた所定のデータを目標物としてとらえることで、例えば「**の近くの〇〇」という発声データについても、「**の近くの」という所定データを得ることで「**」を目標物として用い本来の目的地の「〇〇」を得ることが可能となる。

【0009】また、第3の発明は、ユーザの発声データを記憶する記憶ステップと、前記ユーザの発声データと音声データベース内の音声データとを比較することにより前記音声データの少なくとも一部を解析する第1解析ステップと、前記第1解析ステップで得られた解析データに基づいて前記音声データベースを切り替える切替ステップと、前記記憶ステップで記憶された発声データを読み出し、読み出された発声データと前記切替ステップで切り替えられた音声データベース内の音声データとを比較することにより再解析する第2解析ステップとを有することを特徴とする。

【0010】また、第4の発明は、第3の発明において、前記発声データはナビゲーション用の目的地データであり、前記第2解析ステップにおいて所定のデータが得られた場合に、該データを前記目的地用の目標物として処理する処理ステップをさらに有することを特徴とする。

【0011】

【発明の実施の形態】以下、図面に基づき本発明の実施形態について、ナビゲーションシステムにおける目的地設定を例にとり説明する。

【0012】図1は、本実施形態の構成ブロック図であり、音声認識機能を有するナビゲーションシステムの構成を示す図である。

【0013】マイク10は、ユーザ（車両乗員）の発声を入力して制御部14に供給する。現在位置検出部12は、GPSや車速センサ、方位センサなどで構成され、車両の現在位置を検出して制御部14に供給する。

【0014】制御部14は、具体的にはマイクロコンピュータで構成され、ナビゲーションに必要な各種制御を実行するとともに、マイク10から入力されたユーザの発声データを解析して目的地を設定する。本実施形態では、この制御部14が第1音声解析手段、第2音声解析手段として機能するとともに、音声データベース18の内、解析に使用するデータを切り替える切替手段として機能する。

【0015】音声データベース18は、制御部14でユーザの発声データを解析する際に発声データと比較すべき音声データを記憶するもので、階層構造を有している。制御部14は、適宜この音声データベース18にアクセスして発声データを解析する。音声データベース18は、例えばCD-ROMやDVD等で構成される。

【0016】地図データ記憶部20は、ナビゲーションに必要な地図データ（表示用地図データ及び経路探索用地図データ）を記憶しており、制御部14は検出された現在位置周辺の地図データを地図データ記憶部20から読み出して表示部24に表示し、あるいは発声データを解析することで得られた目的地に至る経路を経路探索用地図データを用いて探索し、推奨経路として表示部24に表示する。もちろん、推奨経路はスピーカから音声で報知してもよい。地図データ記憶部20は、例えばCD-ROMやDVD等で構成される。

【0017】発声データ記憶部22は、マイク10から入力した発声データを記憶するもので、制御部14はこの発声データ記憶部22に記憶された発声データを読み出すことで、ユーザに対して再度の発声を要求することなく発声データの複数回の解析を行うことができる。発声データ記憶部22は、例えば半導体メモリで構成することができる。

【0018】なお、操作部16は、表示部24に表示された地図データのスクロールや、音声によらない手動の目的地設定等の各種入力操作に用いられる。

【0019】図2には、音声データベース18の階層構造が示されている。音声データベースは、全国レベルの認識グラマー辞書、都道府県レベルの認識グラマー辞書、市町村レベルの認識グラマー辞書の3階層から構成されている。なお、「グラマー辞書」とは、制御部14で発声データを解析する際に用いられるグラマー手法における音声データの集合であり、グラマー手法については後述する。全国レベルの認識グラマー辞書には日本全国の主要な地名や名前のデータが記憶され、都道府県レベルの認識グラマー辞書には各都道府県毎に区分されてその都道府県内の地名や名前のデータが記憶され、市町村レベルの認識グラマー辞書には各市町村毎に区分されてその市町村内の地名や名前のデータが記憶されている。

【0020】音声データベース18のどの階層のどのデータを読み出して利用するかは、検出された現在位置及び発声データの解析結果に応じて制御部14が決定する。具体的には、例えば車両の現在位置が静岡県裾野市である場合には、制御部14は音声データベース18内の都道府県レベルの認識グラマー辞書として静岡県を指定し、市町村レベルの認識グラマー辞書として裾野市及びその近隣の市を指定する。そして、車両が移動して車両の現在位置が東京都千代田区となった場合には、制御部14は都道府県レベルの認識グラマー辞書として東京都を指定し、市町村レベルの認識グラマー辞書として千代田区及びその近隣の区を指定する。現在位置に応じた音声データベースを指定する利点は、現在位置周辺を目的地として発声した場合に、迅速に解析して認識できる点である。また、制御部14は、発声データの解析結果に応じて音声データベース18の利用データを切り替え

る。例えば、発声データの解析により三島市が対象となっていることが分かれば、市町村レベルの認識グラマー辞書を三島市に切り替えて解析を続行する。

【0021】図3には、本実施形態における音声認識の処理フローチャートが示されている。まず、ユーザが発声して目的地を入力する(S101)。発声の形態としては、例えば「みしましの**しょうてん(三島市の**商店)」や、「みしまえきのちかくのちゅうしゃじょう(三島駅の近くの駐車場に行きたい)」などである。マイク10から入力されたこれらの発声データは、発声データ記憶部22に記憶されるとともに(S102)、制御部14は入力した発声データをグラマー手法を用いて解析する(S103)。

【0022】ここで、グラマー手法について説明する。グラマー手法とは、認識させたい語の並びを予め定義して認識させる手法であり、例えば文の並びを<a><c>とし、<a>として「今日は」あるいは「明日は」あるいは「あさっては」を候補とし、として「天気は」あるいは「天気が」を候補とし、<c>として「いい」あるいは「悪い」を候補として、発声データ「今日は天気がいい」や「今日は天気が悪い」や「明日は天気がいい」などを認識するものである。本実施形態では、目的地認識を行うために、以下のような語の並び(フレーズと称する)を用いている。

【0023】基本フレーズ1=<地名><end>?
基本フレーズ2=<NULL>?<名前><end>?
基本フレーズ3=<地名>の<NULL>?<名前><end>?

基本フレーズ4=<地名>の<名前>の<NULL>?
<名前><end>?

基本フレーズ5=<名前>の<NULL>?<名前><end>?

基本フレーズ6=<地名><方向><end>?

但し、<地名>は住所あるいはエリアを表す語句で、住所は「しずおか」や「しずおかけん」等であり、エリアは「いず」や「ほうそう」等である。また、<end>は文の終わりを示す語句で、「にいきたい」「へいきたい」「にとめたい」「にかえる」「にいく」「たのむ」「まで」「までのみ」「へ」等である。<NULL>は範囲や程度を表す語句で、「ちかくの」「しゅうへんの」「いちばん」「いちばんちかくの」「ちかい」「やすい」「うまい」「おいしい」「いつもの」「そばの」「こちら」等である。この<NULL>データは目標物を設定する際に必要となるデータでもある。<名前>は名称や施設を表す語句で「**えき」「**ちゅうしゃじょう」「**ごるふじょう」「**こうえん」「**いんたーちえんじ」「**びょういん」「**みなと」「**かわ」「**かんこうち」「**おんせん」等である。<方向>は「**ほうめん」等である。また、<の後の?>は、その<>の語句が必須ではなく、なくて

も良いことを示す。したがって、基本フレーズ1には「しずおかにいきたい(静岡に行きたい)」の他、「しずおか」も含まれる。上述の例における「みしましの**商店(三島市の**商店)」は基本フレーズ3に該当し、「みしまえきのちかくのちゅうしゃじょう(三島駅の近くの駐車場)」は基本フレーズ5に該当する。

【0024】このようなグラマー手法を用いてユーザの発声データを解析すると、少なくとも一部は解析できるものの、残りのデータ(特に発声の後半部分のデータ)が解析できない場合が生じる。具体的には、上述の発声データ「みしましの**しょうてん(三島市の**商店)」を解析した場合、「みしまし(三島市)」は全国レベルの認識グラマー辞書には存在して解析可能であるが、「**しょうてん(**商店)」という名前は市町村レベルの認識グラマー辞書でなければ解析できず、しかも市町村レベルの認識グラマー辞書が三島市以外の市町村に指定されている場合(例えば車両の現在位置が裾野市である場合には、市町村レベルの認識グラマー辞書のデフォルト値は裾野市)には発声データを解析できない。そこで、解析して得られた結果を用いて音声データベース18のグラマー辞書を切り替える(S104)。上述の場合、「みしまし(三島市)」が得られているので、市町村レベルの認識グラマー辞書を三島市用のデータに切り替える。

【0025】音声データベース18を切り替えた後、S102の処理で発声データ記憶部22に記憶した発声データを読み出し、再度解析する(S105)。このとき、市町村レベルの認識グラマー辞書は三島市用のデータとなっているため、発声データの内の「**しょうてん(**商店)」を解析することができる。発声データのすべての解析が終了すると、制御部14は解析結果を用いて地図データから目的地を検索する(S106)。この例では、三島市用の地図データを読み出して**商店を検索することになる。

【0026】一方、発声データが「みしまえきのちかくのちゅうしゃじょうに行きたい(三島駅の近くの駐車場に行きたい)」の場合でも、同様にしてS103で解析を行い、アクティブな認識グラマー辞書(この場合は全国レベル)がヒットして「みしまえき(三島駅)」「ちかくの(近くの)」「ちゅうしゃじょう(駐車場)」を解析することができる。そして、市町村レベルの認識グラマー辞書を三島市用のデータに切替え(S104)、発声データ記憶部22に記憶された発声データを読み出して再度解析を行う(S105)。なお、この例の場合では、1回目の解析で全ての発声データを解析できるので、2回目の解析結果は1回目の解析結果と同一である。もちろん、発声データが「みしまえきのちかくの**ちゅうしゃじょうに行きたい(三島駅の近くの**駐車場に行きたい)」である場合には、1回目の解析では「**ちゅうしゃじょう(**駐車場)」の部分は解析

不能となり、音声データベースを切り替えた後の2回目の解析で「* * ちゅうしゃじょう」の部分も解析できることになる。そして、〈NULL〉データである「近くの」が存在するため、制御部14は〈NULL〉データの前に存在する〈名前〉データの解析結果を目標物として処理し、地図データ上でこの目標物（三島駅）の座標（X、Y）から近い順に駐車場を検索する（S106）。

【0027】このように、本実施形態ではユーザの発声データを記憶しておき、1回目の解析で解析できなかった場合でも音声データベースを自動的に切り替えて再度解析を行うので、音声認識の精度が向上するとともに、ユーザは1回の発声で目的地を設定することができる。

【0028】また、本実施形態では、〈NULL〉データが存在する場合には、その前のデータを目標物とみなして地図データから検索し、その目標物周辺の地図データを検索することで本来の目的地を検索することができるので、ユーザは自然な発声で所望の目的地を容易に設定することが可能となる。

【0029】また、本実施形態において、同音異義語が存在する場合には、認識率を向上させるためにユーザに対してより多くの情報を求めることが好適である。例えば、ユーザが「とよた」と発声した場合、スピーカから「とよたしですか、とよたちょうですか」と問い合わせる等である。

【0030】さらに、発声データの解析を行う場合、得

られたデータの種類の表すアノテーションを付与し、地図データベースの検索を容易とする等の技術は当然ながら本実施形態においても用いることができる。例えば、地名の中の県名にはアノテーションとして数字の11、市名には数字の13、エリアには42を付与し、名前にはアノテーションとして数字の32を付与する等である。この場合、〈NULL〉データの内の「ちかい」や「ちかくの」にはアノテーションとして数字を付与する（例えば91）一方、〈NULL〉データの内の「うまい」や「やすい」にはアノテーションを付与しないのが好適である。これらは目的地の設定（地図データの検索）にとって不要な語句だからである。

【0031】

【発明の効果】以上説明したように、本発明によればユーザの発声の負担を軽減してより簡易に所望のデータ、例えばナビゲーション用の目的地を音声で設定することができる。

【図面の簡単な説明】

【図1】 実施形態の構成ブロック図である。

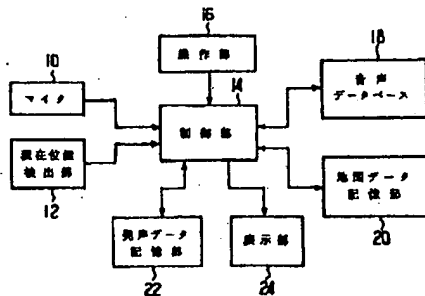
【図2】 実施形態の音声データベースの構造を示す説明図である。

【図3】 実施形態の処理フローチャートである。

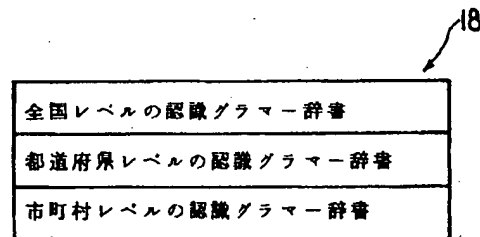
【符号の説明】

10 マイク、12 現在位置検出部、14 制御部、16 操作部、18 音声データベース、20 地図データ記憶部、22 発声データ記憶部、24 表示部。

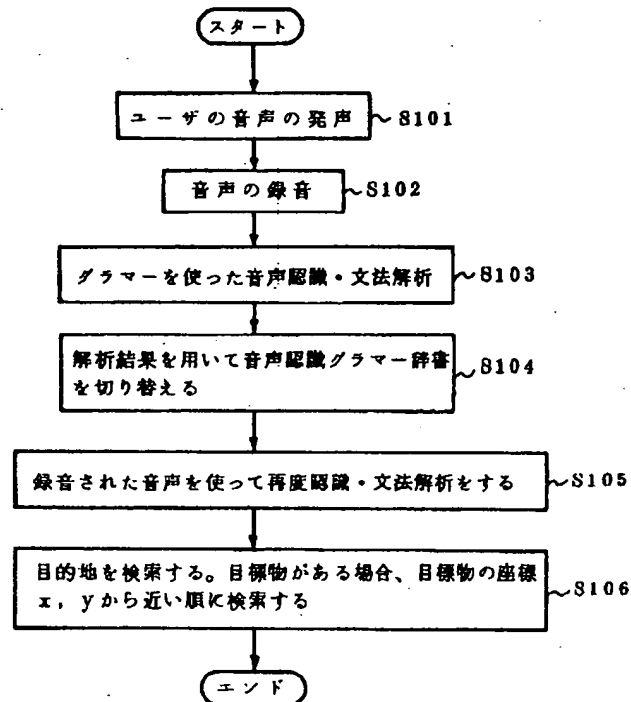
【図1】



【図2】



【図3】



フロントページの続き

(51)Int.Cl.⁷
G 0 9 B 29/10

識別記号

F I
G 0 9 B 29/10

テーム(参考)

Z

Fターム(参考) 2C032 HB06 HC16 HD16
2F029 AA02 AB01 AB07 AB09 AC01
AC02 AC04 AC18
5D015 HH13 HH16 KK02 LL10
5H180 AA01 BB04 BB13 CC12 CC27
FF04 FF05 FF22 FF25 FF32
FF33

(English Translation)

(19) Japan Patent Office

(12) Laid-Open Patent Publication

(11) Laid-Open Patent Publication No.: P2000-181485A

(43) Date of Publication: June 30, 2000

(21) Application No. : 10-354995

(22) Filing Date: December 14, 1998

(71) Applicant: Toyota Motor Corporation

(54) [Title of the Invention] Device and Method for Voice Recognition

(57) [Abstract]

[Problem to be Solved] To lighten the burden of speaking on a user when a destination for navigation is vocally set.

[Solution] The user inputs destination data for navigation through a microphone 10. A control unit 14 stores the input spoken data in a spoken data storage unit 22 and analyzes at least part of the input spoken data by using a voice database 18. The voice database 18 used for the analysis is changed by using the analysis result, and the spoken data stored in

the spoken data storage part 22 are read out to take an analysis again. Thus, the stored spoken data are used, so the user speaks only once to set the destination with high precision.

[Claims for Patent]

[Claim 1] A voice recognition device characterized by comprising:

spoken data storage means for storing spoken data of a user;

first voice analysis means for analyzing at least a portion of the spoken data by comparing the spoken data and voice data in a voice database;

changing means for changing the voice database based upon analysis data obtained by the first voice analysis means; and

second voice analysis means for re-analyzing the spoken data by reading out the spoken data stored in the spoken data storage means, and comparing the spoken data with voice data in the voice database changed by the changing means.

[Claim 2] The voice recognition device according to claim 1, characterized in that the spoken data is destination data for navigation, and further comprising:

means for processing predetermined data as a landmark for a destination when the

predetermined data is obtained by the second voice analysis means.

[Claim 3] A voice recognition method characterized by comprising the steps of:

storing spoken data of a user;

a first analysis for analyzing at least a portion of voice data by comparing the spoken data of the user and the voice data in a voice database;

changing the voice database based upon analysis data obtained by the first voice analysis; and

a second analysis for re-analyzing by reading out the spoken data stored in the storing step, and comparing the spoken data read out with voice data in the voice database changed in the changing step.

[Claim 4] The voice recognition method according to claim 3, characterized in that the spoken data is destination data for navigation, and further comprising the step of:

processing predetermined data as a landmark for a destination when the predetermined data is obtained by the second analysis.

[Detailed Description of the Invention]

[0001]

[Technical Field of the Invention] The present invention relates to a device and method for voice recognition, and in particular, relates to voice recognition for setting a destination in a navigation system.

[0002]

[Related Art] Art has been proposed in the past for executing various processing by voice, such as setting a destination, in a navigation system. In such art, an important issue is how fast and accurate a voice spoken by a user can be recognized. Voice recognition is usually performed by comparing user spoken data and voice data in a voice database prepared in advance. A hierarchical voice database is often used.

[0003] For example, an art is disclosed in Japanese Patent Laid-Open Publication No. 10-62199, in which a voice database is separated into three layers. Layer 1 stores facility names with location information, and facility category names without location information. Layer 2 stores facility names with location information and prefecture names without location information, which correspond to the category names stored in layer 1. Layer 3 stores facility names with location information corresponding to the prefectures in layer 2. Voice recognition is performed by changing sequential layers in accordance with user

spoken data.

[0004]

[Problem to be Solved by the Invention] In the above related art, the layers of the voice database are changed each time the user speaks. Therefore, if the user wants to set the navigation destination as XX department store in a department store category, he or she must repeatedly say [facility] → [department store] → [XX department store] in that order. The destination could not be set by saying a natural phrase once, such as “I want to go to a department store, the XX department store”.

[0005] Furthermore, some users may want to set the destination based on a certain landmark object, such as a parking lot near the XX store, for example. However, it was impossible for the related art to recognize a destination setting based on a landmark object in this manner.

[0006] In light of the problems found in the above related art, it is an object of the present invention to provide a device and method capable of more easily setting desired data by voice and lightening the burden of speaking on a user.

[0007]

[Means for Solving the Problem] In order to achieve the above object, a first invention is characterized by including spoken data storage means for storing spoken data of a user; first voice analysis means for analyzing at least a portion of the spoken data by comparing the spoken data and voice data in a voice database; changing means for changing the voice database based upon analysis data obtained by the first voice analysis means; and second voice analysis means for re-analyzing the spoken data by reading out the spoken data stored in the spoken data storage means, and comparing the spoken data with voice data in the voice database changed by the changing means. Voice recognition can be surely performed after the user speaks once by analysis with the first voice analysis means, and further analysis again with the second voice analysis means after reading out stored voice data (with the voice database changed and optimized for re-analysis). It should be noted that the first voice analysis means and the second voice analysis means do not have to be separate, and both functions may also be achieved by the same means.

[0008] A second invention according to the first invention is characterized in that the spoken data is destination data for navigation, and further includes means for processing predetermined data as a landmark for a destination when the predetermined data is obtained

by the second voice analysis means. By designating predetermined data obtained in a voice analysis as a landmark, for example, in spoken data “XX near YY”, it is possible to obtain the actual destination “XX” after obtaining the predetermined data “near YY” using “YY” as a landmark.

[0009] A third invention is characterized by including the steps of: storing spoken data of a user; a first analysis for analyzing at least a portion of voice data by comparing the spoken data of the user and the voice data in a voice database; changing the voice database based upon analysis data obtained by the first voice analysis; and a second analysis for re-analyzing by reading out the spoken data stored in the storing step, and comparing the spoken data read out with voice data in the voice database changed in the changing step.

[0010] A fourth invention according to the third invention is characterized in that the spoken data is destination data for navigation, and further includes the step of processing predetermined data as a landmark for a destination when the predetermined data is obtained by the second analysis.

[0011]

[Embodiments of the Invention] Hereinafter, embodiments of the present invention will

be described with reference to the accompanying drawings in an example of destination setting in a navigation system.

[0012] FIG. 1 is a structural block diagram of the present embodiment, and a drawing showing the structure of a navigation system with a voice recognition function.

[0013] A voice of a user (vehicle passenger) is input into a microphone 10, which is then supplied to a control unit 14. A present location detecting unit 12 includes a GPS, vehicle speed and direction sensors. The present location detecting unit 12 detects the present location of a vehicle, which it supplies to the control unit 14.

[0014] The control unit 14 specifically includes a microcomputer. In addition to executing various controls required for navigation, the control unit 14 analyzes user spoken data input from the microphone 10 to set a destination. In the present embodiment, the control unit 14 functions as first voice analysis means and second voice analysis means. Moreover, the control unit 14 functions as changing means for changing data used for analysis inside a voice database 18.

[0015] The voice database 18 has a hierarchical structure, and stores voice data for comparison with spoken data when the control unit 14 analyzes user spoken data. The

control unit 14 accesses the voice database 18 as necessary to analyze spoken data. In addition, the voice database 18 may be provided with a CD-ROM, DVD or the like.

[0016] A map data storage unit 20 stores map data (map data for display and map data for route searches) required for navigation. The control unit 14 reads out map data in the vicinity of the detected present location from the map data storage unit 20, which is displayed on a display unit 24. Alternately, using map data for route searches, the control unit 14 searches for a route to a destination that was obtained by analyzing spoken data. The recommended route is then displayed on the display unit 24. Naturally, the recommended route may be reported by voice from a speaker. In addition, the map data storage unit 20 may be provided with a CD-ROM, DVD or the like.

[0017] A spoken data storage unit 22 stores spoken data input from the microphone 10. By reading out spoken data stored in the spoken data storage unit 22, the control unit 14 can perform multiple analyses of spoken data without requiring additional speaking by the user. In addition, the spoken data storage unit 22 may be provided with a semiconductor memory.

[0018] It should be noted that an operating unit 16 is used for various input operations,

such as scrolling map data displayed on the display unit 24, and manually setting a destination without speaking.

[0019] FIG. 2 shows the hierarchical structure of the voice database 18. The voice database has three layers: a national-level recognition grammar dictionary, a prefectural-level recognition grammar dictionary, and a municipal-level recognition grammar dictionary. Note that a “grammar dictionary” is a collection of voice data for a grammar method used in the analysis of spoken data by the control unit 14, and the grammar method will be described later. The national-level recognition grammar dictionary stores data for main places and names for all of Japan; the prefectural-level recognition grammar dictionary is classified by each prefecture, and stores data for places and names within prefectures; and the municipal-level recognition grammar dictionary is classified by each city, and stores data for places and names within cities.

[0020] The control unit 14 decides what data from which layer of the voice database 18 is read out and used, based upon the detected present location and the analysis result of spoken data. More specifically, if the present location of the vehicle is the city of Susono in Shizuoka Prefecture, for example, the control unit 14 specifies Shizuoka Prefecture as

the prefectural-level recognition grammar dictionary, and Susono and neighboring cities as the municipal-level recognition grammar dictionary within the voice database 18. If the vehicle then moves so that its present location is Chiyoda ward in Tokyo, the control unit 14 specifies Tokyo as the prefectural-level recognition grammar dictionary, and Chiyoda ward and neighboring wards as the municipal-level recognition grammar dictionary. One advantage of specifying the voice database depending on the present location is a quick analysis and recognition ability when the vicinity of the present location is spoken as the destination. Furthermore, the control unit 14 changes the data used in the voice database 18 depending on the analysis result of spoken data. For example, once it is known that the city of Mishima is a landmark according to the analysis result of spoken data, the analysis is continued after changing the municipal-level recognition grammar dictionary to Mishima.

[0021] FIG. 3 shows a processing flowchart of voice recognition according to the present embodiment. First, the user speaks to input a destination (S101). The user may say something such as, "XX store in Mishima City" or "I want to go to the parking lot near Mishima Station". This spoken data input from the microphone 10 is stored in the spoken

data storage unit 22 (S102), and the control unit 14 analyzes the input spoken data using a grammar method (S103).

[0022] The grammar method is described here. In the grammar method, recognition is carried out through the predefinition of combinations of words to be recognized. For example, a sentence combination consists of <a><c>, where the candidates for <a> are “today”, “tomorrow” and “the day after tomorrow”; the candidates for are “the weather is” or “the weather will be”; and the candidates for <c> are “good” or “bad”. In this case, sentences such as “today the weather is good”, “today the weather is bad”, and “tomorrow the weather will be good” are recognized. Word combinations (called phrases) such as the following are used in the present embodiment to recognize destinations.

[0023] Basic phrase 1 = <end><place>?

Basic phrase 2 = <end><name>?<NULL>?

Basic phrase 3 = <end><name>?<NULL> in <place>?

Basic phrase 4 = <end><name>?<NULL> in <name> in <place>?

Basic phrase 5 = <end><name>?<NULL> in <name>?

Basic phrase 6 = <end><direction><place>?

Note that <place> is a phrase that represents an address or area, where the address may be “Shizuoka” or “Shizuoka Prefecture” or the like, and the area may be “Izu” or “Bousou” or the like. <end> is a phrase that indicates the end of a sentence, such as “I want to go to”, “take me to”, “I want to stop at”, “return to”, “go to”, “please”, “to”, “please go over to”, and “over to”. <NULL> is a phrase that represents a range or degree, such as “near”, “around”, “first”, “closest to”, “close to”, “cheap”, “good food”, “delicious”, “the usual”, “close by”, and “around here”. <NULL> data is also data required when setting a landmark. <name> is a phrase that represents a name or a facility. Examples of <name> include “XX station”, “XX parking lot”, “XX golf course”, “XX park”, “XX interchange”, “XX hospital”, “XX harbor”, “XX river”, “XX tourist resort”, and “XX hot springs”. <direction> includes “toward XX” or the like. A question mark after a bracket indicates that the phrase in the brackets is not essential and may be ignored. Accordingly, the basic phrase 1 includes “Shizuoka” as well as “I want to go to Shizuoka”. The examples mentioned earlier, “XX store in Mishima City” and “the parking lot near Mishima Station” correspond to the basic phrases 3 and 5, respectively.

[0024] By using such a grammar method to analyze user spoken data, it is possible to

analyze at least a portion of the spoken data, although the rest of the spoken data (especially the latter half of spoken data) may not permit analysis. More specifically, when analyzing the above spoken data “XX store in Mishima City”, the word “Mishima City” is in the national-level recognition grammar dictionary and can be analyzed. However, the name “XX store” cannot be analyzed unless the municipal-level recognition grammar dictionary is used; moreover, the spoken data cannot be analyzed if a city other than Mishima City is specified in the municipal-level recognition grammar dictionary (i.e., if the present location of the vehicle is Susono City, the default value of the municipal-level recognition grammar dictionary is Susono City). Thus, the grammar dictionary of the voice database 18 is changed using the result obtained from analysis (S104). Since “Mishima City” was obtained in the above example, the municipal-level recognition grammar dictionary is changed to data for Mishima City.

[0025] After changing the voice database 18, the spoken data stored in the spoken data storage unit 22 in processing at S102 is read out and analyzed again (S105). At this time, the municipal-level recognition grammar dictionary is set to data for Mishima City, thus, the words “XX store” in the spoken data can be analyzed. Once the spoken data is

completely analyzed, the control unit 14 searches for the destination from map data using the analysis result (S106). In this example, map data for Mishima City is read out to search for XX store.

[0026] Meanwhile, spoken data consisting of “I want to go to the parking lot near Mishima Station” can also be analyzed in the same manner in S103. In this case, the active recognition grammar dictionary (national-level in this case) can pick up and analyze the words “Mishima Station”, “near”, and “parking lot”. Thereafter, the municipal-level recognition grammar dictionary is changed to data for Mishima City (S104). Next, the spoken data stored in the spoken data storage unit 22 is read out and analyzed again (S105).

It should be noted that in the case of this example, the second analysis result is the same as the first analysis result because it was possible to analyze all the spoken data in the first analysis. Naturally, the words “XX parking lot” in the case of spoken data consisting of “I want to go to the XX parking lot near Mishima Station” cannot be analyzed in the first analysis. The “XX parking lot” portion can be analyzed in the second analysis after changing the voice database. Since there is also the <NULL> data “near”, the control unit 14 processes the analysis result of the <name> data before the <NULL> data as a landmark,

and searches for parking lots starting from those closest to the coordinates (X, Y) of the landmark (Mishima Station) in the map data (S106).

[0027] As described above, in the present embodiment, user spoken data is stored such that the voice database is automatically changed for re-analysis even when an analysis could not be performed the first time. Thus, the precision of voice recognition is improved, in addition to allowing the user to set the destination after speaking only once.

[0028] In the present embodiment, if there is <NULL> data, data before <NULL> data is considered a landmark, based upon which a search is performed in map data. Hence, the actual destination can be found by searching map data in the vicinity of the landmark, thereby allowing the user to easily set a desired destination using natural vocalization.

[0029] Furthermore, in the present embodiment, requesting more information from the user is suitable for improving the rate of recognition in the case of homonyms. For example, if the user says "Toyota", a question such as "Toyota City or Toyota Town?" may be output from a speaker.

[0030] Moreover, when analyzing spoken data, art such as assigning annotation representing the types of obtained data to facilitate map database searches may naturally be

used in the present embodiment. For example, the numeral 11 can be assigned as annotation for a prefecture, 13 for a city name, 42 for an area, and 32 for a name in a place. In this case, while a numeral (i.e., 91) is assigned as annotation for <NULL> data such as “close to” and “near”, no annotation should be assigned for <NULL> data such as “cheap” and “good food”. This is because such words are not required for setting a destination (searching map data).

[0031]

[Effect of the Invention] As described above, according to the present invention, it is possible to set desired data, such as a destination for navigation, more easily by voice and lighten the burden of speaking on a user.

[Brief Description of the Drawings]

[FIG. 1] FIG. 1 is a structural block diagram of an embodiment.

[FIG.2] FIG. 2 is an explanatory drawing of a voice database structure of the embodiment.

[FIG. 3] FIG. 3 is a processing flowchart of the embodiment.

[Description of Reference Numerals]

10 MICROPHONE

- 12 PRESENT LOCATION DETECTING UNIT
- 14 CONTROL UNIT
- 16 OPERATING UNIT
- 18 VOICE DATABASE
- 20 MAP DATA STORAGE UNIT
- 22 SPOKEN DATA STORAGE UNIT
- 24 DISPLAY UNIT

[FIG. 1]

10/MICROPHONE

12/PRESENT LOCATION DETECTING UNIT

14/CONTROL UNIT

16/OPERATING UNIT

18/VOICE DATABASE

20/MAP DATA STORAGE UNIT

22/SPOKEN DATA STORAGE UNIT

24/DISPLAY UNIT

[FIG. 2]

NATIONAL-LEVEL RECOGNITION GRAMMAR DICTIONARY

PREFECTURAL-LEVEL RECOGNITION GRAMMAR DICTIONARY

MUNICIPAL-LEVEL RECOGNITION GRAMMAR DICTIONARY

[FIG. 3]

START

S101/USER SPEAKS

S102/RECORD USER VOICE

S103/USE GRAMMAR FOR VOICE RECOGNITION / GRAMMAR ANALYSIS

S104/CHANGE VOICE RECOGNITION GRAMMAR DICTIONARY USING
ANALYSIS RESULT

S105/USE RECORDED VOICE FOR RE-RECOGNITION / GRAMMAR RE-ANALYSIS

S106/SEARCH FOR DESTINATION. IF LANDMARK USED, SEARCH IN ORDER
STARTING FROM x, y COORDINATES OF LANDMARK

END